Effective Selection of Targeted Advertisements for Vehicular Users

Gil Einziger
Politecnico di Torino, Italy

Carla Fabiana Chiasserini
Politecnico di Torino, Italy

Francesco Malandrino
Politecnico di Torino, Italy

ABSTRACT

This paper focuses on targeted advertising for vehicular users, where users receive advertisements (ads) from roadside units and the vehicle onboard system displays only ads that are relevant to the user. A broker broadcasts ads and is paid by advertisers based on the number of vehicles that displayed each ad. The problem we study is the following: given that the broker can broadcast a limited number of ads, what is the strategy for ad selection that maximizes the broker's revenue? We first identify the conflict existing between users' interests and broker's revenue as a critical feature of this scenario, which may dramatically reduce the broker's revenue. Then, given the problem complexity, we propose Volfied, an algorithm that solves this conflict, allows for near-optimal broker's revenue and has very limited computational complexity. Our results show that Volfied increases the broker's revenue by up to 70% with respect to state-of-the-art alternatives.

Keywords

Performance evaluation; algorithms; services for mobile users

1. INTRODUCTION

As mobile devices outnumber TV sets and desktop PCs [1], advertisers have recently rushed to this media. Mobile advertising is therefore growing globally at a rapid rate, involving an increasing variety of mobile devices [2]. This paper focuses on targeted advertising for vehicular users, although our problem formulation and solution can be easily extended to the case of pedestrian users. Vehicular users are indeed expected to represent a large portion of the mobile users population in a few years, and several types of business (shops, restaurants, touristic attractions) are interested in advertising their products and services through, e.g., onboard devices as an alternative to static advertisement billboards.

In particular, targeted advertising (such as Google AdWords and Ink TAD) aims at displaying advertisements (ads) only to interested users, by analyzing their online profiles or behaviors [3]. This approach has been shown to be very beneficial for both advertisers and users [4]. On one hand, the probability of an ad being effective increases significantly with the user interest in the product or service. On the other, users are not bothered with irrelevant information. This is particularly important in a mobile scenario, where users cannot be exposed to too many ads at the same time, due to their reduced screen size and attention span.

In accordance with Internet advertising systems [5, 6], our work addresses a scenario where advertisers sign a contract with a broker (e.g., an advertising platform) to display an ad to interested users. The user exposure to the ad is referred to as an *impression* [7]. Each ad may have a different value depending on the advertised product/service, and the broker is paid by advertisers based on both the ad value and the number of ad impressions. In our vehicular environment, the broker delivers ads by broadcasting them through roadside units (e.g., base stations or APs), and the number of broadcast ads is constrained by bandwidth limitations. Vehicular users passing by a roadside unit (RSU) will receive the transmitted ads; however, as mentioned, the onboard system will display only ads that are relevant to the user. The actual number of impressions can then be reported to the broker by vehicles, periodically (e.g., once a day) and in a secure manner.

Under this framework, our goal is to solve the problem of real-time ads selection that the broker should perform in order to maximize its revenue while meeting the constraint on the number of ads that can be broadcast. The broker's revenue is defined as the number of ad impressions, each weighted by the value of the displayed ad. To find the best ad selection strategy, the broker exploits information on radio coverage and users' interests. The former can be obtained through RSUs, the latter can be obtained from the user profile upon service subscription.

The main contributions of our work are as follows.

(i) We present a system model that jointly captures ads features and users' interests. We show that when ads are broadcast by the RSUs and the users can choose which ads to display, there exists a conflict between the broker and users' interests. This implies that increasing the number of broadcast ads may actually reduce the broker's revenue. We then formulate the following optimization problem: which is the best set of ads that the broker should select to maximize its revenue? Due to its complexity, we design a greedy algorithm, named after the computer game Volfied. Our algorithm solves the above conflict and provides an efficient solution.

(ii) In order to further speed up the ad selection procedure, we propose a technique to reduce the number of ads to process. By applying Volfied to such simplified system representation, we ensure a swifter on-line ad selection, with negligible performance loss.

(iii) Finally, we evaluate Volfied in a realistic vehicular environment and show that its performance is nearly optimal in a small-scale scenario. In a large-scale scenario, Volfied is compared to other heuristics, such as Top-k and Random, and it is shown to

MSWiM '16, November 13-17, 2016, Malta, Malta

DOI: http://dx.doi.org/10.1145/2988287.2989136

Table 1: Notation

Symbol	Description
$\mathcal{A} = \{a\}$	Set of advertisements
$\mathcal{V} = \{v\}$	Set of vehicular users
$\mathcal{U} = \{u\}$	Set of RSUs
$\mathcal{A}_\epsilon^{(M)}$	M-sparse approximation of set $\mathcal{A}$
$D(\cdot,\cdot)$	Distance between users' interests and/or ads
K	No. of ads each RSU can broadcast in one time step
M	No. of ads each vehicle can display in one time step
$r(a, u)$	Value of ad a under the coverage of RSU u
$R(a, u)$	Estimated total revenue for ad $a \in \mathcal{A}$ at RSU u

increase the broker's revenue by up to 70% and the generated impressions by up to 50%.

Our model and problem formulation share some common elements with previous work, as discussed in Sec. 2. However, to the best of our knowledge, our work is the first to present a unified model capturing the main aspects of mobile advertising, to identify the conflict issue between users' interests and broker's revenue, and to propose a highly efficient algorithm for ad selection that meets the system constraints.

2. RELATED WORK

Advertisement scheduling is an important problem that can be studied from multiple perspectives, as it provides different challenges for advertisers, brokers and users. Existing works typically employ machine learning [8] or game theory approaches [9], and either focus on ad pricing or attempt to maximize social welfare. For example, [10, 11] suggest methods for advertisers to configure their ads features and generation speed to maximize ads visibility. Similarly, [12] suggests ad scheduling techniques to maximize revenue over a shared medium. This approach requires advertisers to make complex decisions, which are not practical in a vehicular environment. Other approaches aim at maximizing social welfare [13] in ad auctions, or address social influence in on-line advertising [6]. Another research direction is to treat selection of on-line ads as an optimization problem [7, 14], which however cannot scale to large systems due to solution complexity. The work in [15] addresses a scenario and problem similar to ours, albeit with a simpler display policy at the vehicle level. The strategy in [15] is radically different from ours: decisions are made by solving an ILP optimization problem, which would be too computationally intensive in our scenario. Finally, several works have addressed privacy in targeted advertising (see, e.g., [5]) – a relevant issue that, however, is out of the scope of this work.

3. SYSTEM MODEL AND PROBLEM FORMULATION

We consider that advertisers rely on an ad platform, called *broker*, to deliver targeted ads to vehicular users. Each user has a profile from which the user's interests can be deduced. The delivery of ads takes place through an infrastructure-based vehicular network, composed of vehicles and RSUs. Through such a network, the broker can disseminate ads to vehicular users.

For simplicity, we consider that time is divided into discrete steps. Vehicles receive ads broadcast by the RSUs under whose coverage they pass, and display the ones that are relevant (i.e., of interest) to the user. At each step, vehicles can display a limited number of ads, M, and ads are not cached. Also, a vehicle displays an ad at most once[1];The number of ads that can be broadcast by RSUs is constrained by bandwidth and cost limitations. We denote the maximum number of ads that each RSU can broadcast in one time step by K.

It is the broker's job to select the sets of ads to be broadcast at each time step by each RSU. In order to make this decision, the broker can use the following information:
(i) users' interests and preferences;
(ii) the vehicles that are currently under coverage by RSUs;
(iii) the ads that have been broadcast by RSUs in the past;
(iv) the RSUs visited by vehicles in the past.

Information about users' interests can be provided by the users themselves upon subscription to the service, or obtained through nowadays-common profiling techniques. Information about the presence of vehicles under RSUs coverage can be obtained from the RSUs themselves, by exploiting the beacons vehicles periodically transmit [16]; in our performance evaluation, we also study how errors in acquiring such information affect the performance.

Model entities. The main entities we need to model are: (i) *vehicles* (also referred as users), $v \in \mathcal{V}$, and (ii) *ads* $a \in \mathcal{A}$. Each ad a has an associated *value*, $r(a, u)$. Having RSU-specific ad values allows us to model both *local* ads, which are worthless at RSUs out of their target location, and *global* ones, whose value is constant at all RSUs. Every time ad a is displayed to a vehicle, the onboard platform notifies the broker, which gets a *revenue* equal to the ad value. Thus the broker's total revenue is given by the number of impressions, each weighted by the value of the displayed ad.

The content of ads and the interest of vehicles are both described in terms of *features* $\vec{f} \in \mathbb{F}$. Therefore, both ads and vehicles can be mapped onto points in an n-dimensional *feature space*, $\mathbb{F} \subseteq \mathbb{R}^n$, where n is the number of features.

Distance and relevance. We can define the *distance* between two points $\vec{f_1}, \vec{f_2} \in \mathbb{F}$ (either ads or vehicular users) as:

$$D(\vec{f_1}, \vec{f_2}) = \left\| \vec{f_1} - \vec{f_2} \right\|_0 . \tag{1}$$

If both $\vec{f_1}$ and $\vec{f_2}$ are vehicles, the distance defined in (1) expresses how similar their interests are. If both are ads, (1) conveys how similar the ads themselves, and their potential audience, are. Finally, if $\vec{f_1}$ is an ad $a \in \mathcal{A}$ and $\vec{f_2}$ is a vehicle $v \in \mathcal{V}$, the distance $D(a, v)$ represents how *relevant* ad a is to vehicle v.

We also define a *relevance threshold* $D_{\max}$: only ads closer than $D_{\max}$ are relevant to a user.

The above notation, along with the one used in the following, is summarized in Tab. 1.

3.1 Problem definition

It is the broker's task to define and enact what we formally call a *selection strategy*: given the set $\mathcal{A}$ of ads, the number of ads that can be broadcast (K) and displayed (M), and the vehicles under RSU coverage, the broker has to select those ads that maximize its revenue. Intuitively, the broker should select ads that will be displayed by many vehicles and have a high value r. The former implies that the selected ads should be relevant *and* new to as many vehicles as possible, but, quite surprisingly, these two conditions are not sufficient to ensure that a broadcast ad is actually displayed by the vehicle. Indeed, the broker decides which ads to broadcast, but vehicles decide which of these ads to display. The aims of these two actors are different and potentially *conflicting*: the broker would aim at selecting ads with high value r, while vehicles display ads based on their relevance to the user. Thus, *whenever the broker*

[1] Ads that generate a revenue when displayed multiple times can be represented by separate elements of $\mathcal{A}$.

can broadcast more ads than vehicles can display (i.e., $K > M$), a conflict between the broker and the users' interests may arise. We remark that indeed $K > M$ in all practical cases, and that, as highlighted in the example below, conflicts do not only waste radio resources, but they can also severely reduce the broker's revenue.

Example. Consider a toy case with one RSU ($\mathcal{U} = \{u\}$), one vehicle ($\mathcal{V} = \{v\}$) and two ads ($\mathcal{A} = \{a_1, a_2\}$). Assume: $r(a_1) = 10$, $r(a_2) = 1$, $D(a_1, v) = 0.1$ and $D(a_2, v) = 0.05$. Also, let us focus on one time step and assume $M = 1$, i.e., the vehicle can display only one ad, and $D_{\max} = 0.15$. First, consider $K = 1$, i.e., the RSU can transmit only one ad, and that the RSU sends a_1. Then the vehicle will display a_1 and the broker will earn $r(a_1, u) = 10$. Now, assume $K = 2$ and that the RSU sends a_1 and a_2: one would expect that by sending more ads, the broker would earn *at least* the same revenue. However, owing to the fact that $M = 1$, vehicle v will disregard a_1 and only display a_2, being the most relevant to itself. Thus, the broker's revenue will be $r(a_2, u) = 1$.

In light of this, we introduce the following definition.

DEFINITION 3.1 (CONFLICT-FREE SET). *A set of selected ads, $\mathcal{S} \subseteq \mathcal{A}$, is conflict free if, for each vehicle $v \in \mathcal{V}$, the set includes at most M ads that are relevant to v.*

3.2 Problem formulation

We now formulate the ad selection problem as an optimization problem. We denote the current time step by t_c, and the set of past and current steps by $\mathcal{T}$. Then the set of binary flags $\chi(u, v, t) \in \{0, 1\}$ express whether RSU u covers vehicle v at time $t \in \mathcal{T}$.

Our formulation involves two binary decision variables: $\beta(a, u, t_c)$ and $\delta(a, v, t_c)$. The former concerns the broker, and it indicates whether an ad a is broadcast by RSU u at the current time step or not. The latter concerns individual vehicles, and it indicates whether ad a is displayed by vehicle v at time t_c. Note that, although the vehicles and the broker make different decisions for different, and indeed conflicting, purposes, we are able to reproduce both decisions in the *same* optimization problem, as laid out next.

Constraints. A vehicle v can display only the relevant ads that it receives from the current RSU, i.e., for any $a \in \mathcal{A}$ and $v \in \mathcal{V}$

$$\delta(a, v, t_c) \le \chi(u, v, t_c)\beta(a, u, t_c)\mathbb{1}_{[D(a,v) \le D_{\max}]}\,, \tag{2}$$

where u is the RSU, and $\mathbb{1}_{[D(a,v) \le D_{\max}]}$ takes 1 if $D(a, v) \le D_{\max}$ and 0 otherwise. Next, vehicles can display at most M ads:

$$\sum_{a \in \mathcal{A}} \delta(a, v, t_c) \le M, \ \ \forall v \in \mathcal{V}. \tag{3}$$

Each ad can be shown at most once by every vehicle:

$$\sum_{t \in \mathcal{T}} \delta(a, v, t) \le 1, \ \ \forall a \in \mathcal{A}, v \in \mathcal{V}. \tag{4}$$

Note that the δ values that refer to previous time steps are input parameters to the problem.

Last, we must make sure that a vehicle v selects the ads to display based on their relevance to itself. In other words, vehicle v will not display an ad a if it receives from the RSU M (or more) ads whose relevance to v is higher than a's:

$$\delta(a, v, t_c) \le \max\left\{0, M - \sum_{\substack{a' \in \mathcal{A}: \\ D(a', v) > D(a, v)}} \left[\chi(u, v, t_c)\beta(a', u, t_c) \left(1 - \sum_{t \in \mathcal{T} \setminus \{t_c\}} \delta(a', v, t)\right)\right]\right\} \ \forall a \in \mathcal{A}, v \in \mathcal{V}. \tag{5}$$

As far as the broker is concerned, the sole constraint is on the maximum number of ads that each RSU can broadcast at a given time step:

$$\sum_{a \in \mathcal{A}} \beta(a, u, t_c) \le K \ \ \forall u \in \mathcal{U}. \tag{6}$$

Objective. Given the above constraints, the broker's objective is to maximize its revenue:

$$\max \sum_{a \in \mathcal{A}} \sum_{v \in \mathcal{V}} \sum_{u \in \mathcal{U}} \delta(a, v, t_c)\chi(u, v, t_c)r(a, u)\,. \tag{7}$$

Discussion. The above formulation has the interesting property of accounting for the way *both* vehicles and broker make decisions. Constraint (5) describes how vehicles will select ads based on the ads' relevance to themselves, while objective (7) represents the broker's aim to maximize its own revenue. Thus, conflicts are accounted for: by solving the optimization problem, the broker will maximize its revenue *subject to* the behavior of the vehicles. On the negative side, the problem complexity prevents its solution in large-scale scenarios. Specifically, the ad selection is a $0 - 1$ knapsack problem with constant weights, whose item values are the outcome of another $0 - 1$ knapsack problem (the selection of the ads to display). Thus, the optimization problem is NP-hard. In light of this, we present below a heuristic approach.

4. ON-LINE DECISION MAKING

Our problem exhibits two main challenges. The first has to do with the *conflict* between the broker's revenue and the user interests, which may significantly impair the broker's revenue. The second is *complexity*, since the set of ads $\mathcal{A}$ is potentially very large, as are the sets of ads relevant to individual vehicles. We address these two challenges separately. First, we propose a way to make *conflict-free* decisions leveraging on the estimated revenue that ads can generate. Then we introduce a *sparse-set* approximation that bounds the complexity of estimating ad revenues. For ease of presentation, we describe our decision-making scheme with reference to one RSU and one time step only, and we drop the RSU and time indices when discussing this scenario. Sec. 4.3 explains how to extend the proposed schemes to the multi-step and multi-RSU cases.

4.1 Conflict-free decisions: Volfied

In order to select a set of ads that maximizes its revenue, the broker has to first estimate the revenue it will get from broadcasting a generic ad a. Let $R(a)$ denote such estimated revenue. $R(a)$ is computed by adding $r(a)$ thereto every time a vehicle v, to which a is relevant, enters the RSU coverage area, and subtracting the same amount when v leaves the coverage area.

Armed with the estimated revenues $R(a)$, the broker applies an ad selection strategy. The most straightforward strategy would be *Top-k*, which selects the K ads with highest estimated revenue $R(a)$. However, Top-k has the major disadvantage of ignoring the fact that vehicles can display at most M ads each, thus it may create conflicts that harm the broker's revenue and waste radio resources on ads that will not be displayed (see Sec. 3.1). To avoid this, we devise a *conflict-free* alternative, called Volfied and presented in Alg. 1.

The objective of Alg. 1 is to identify the set $\mathcal{S} \subseteq \mathcal{A}$ of ads to broadcast, initialized in line 1. Volfied starts by sorting set $\mathcal{A}$ by estimated revenue, in line 2. Then, for each ad a, it checks how many ads are already in $\mathcal{S}$ closer to a than $2D_{\max}$ (line 4). If less than M, a is added to the set of ads to serve, in line 5. The algorithm ends when either all ads have been evaluated, or K ads have been selected (line 6).

Algorithm 1 Conflict-free ad selection: Volfied

Require: $\mathcal{A}, K, M, D_{\max}, R(a)$
1: $\mathcal{S} \leftarrow \emptyset$
2: **sort** $a \in \mathcal{A}$ **by** $R(a)$ in decreasing order
3: **for all** $a \in \mathcal{A}$ **do**
4: **if** $|\{b \in \mathcal{S} \colon D(a,b) \leq 2D_{\max}\}| < M$ **then**
5: $\mathcal{S} \leftarrow \mathcal{S} \cup \{a\}$
6: **if** $|\mathcal{S}| \geq K$ **then**
7: **break**
return $\mathcal{S}$

Below, we formally prove that Volfied always generates a conflict-free set of ads, i.e., no vehicle gets more than M relevant ads.

Theorem 1. *The set of ads $\mathcal{S}$ selected by Volfied is conflict free.*

PROOF. Consider a set with one ad only; this is clearly conflict free. Then, by construction (line 4), Volfied selects an additional ad only if, for any ad $a \in \mathcal{S}$, there are less than M ads within distance $2D_{\max}$. This implies that, for any vehicle v, $\mathcal{S}$ includes at most M ads relevant to v, i.e., $\mathcal{S}$ remains conflict free. Indeed, due to triangle inequality, for any two ads a and b s.t. $D(a,b) > 2D_{\max}$, we have: $D(a,v) + D(v,b) \geq D(a,b) > 2D_{\max}$, for any vehicle v. That is, given an ad a, which is relevant to v, only ads within distance $2D_{\max}$ from a may be relevant to v too. □

Furthermore, for sake of completeness, we remark that, in the special case $M = K$, Volfied outputs the same ad set as Top-k, and such set maximizes the broker's revenue. In other words, both Volfied and Top-k are optimal when $K = M$. This is because, by selecting the first K top-revenue ads, the condition in line 4 in Alg. 1 is always met (as $K = M$), thus Volfied and Top-k select the same ads. By Theorem 1, the set is conflict free; also it maximizes the broker's revenue since, by construction, it includes the K ads with top estimated revenue $R(a)$. Next, we show that Volfied has linear complexity in the size of the ads set.

Theorem 2. *The worst case runtime complexity of Volfied is $O(|\mathcal{A}| \cdot K)$.*

PROOF. From Algorithm 1, one can see that the loop in line 3 iterates over all the ads $a \in \mathcal{A}$, thus in the worst case all ads in $\mathcal{A}$ are processed. In line 4, we compare each ad against all previously selected ads, which are at most $K - 1$. The operations in the remaining lines have complexity $O(1)$ and thus the overall complexity of Algorithm 1 is $O(|\mathcal{A}| \cdot K)$. □

Finally, we remark that Volfied relies on the estimated revenues, i.e., the $R(a)$ values. Such estimates need to be refreshed every time a vehicle enters or exits the coverage area of an RSU. Every update has a linear cost in the number of ads, as shown below.

Property 1. *The worst case complexity of updating the revenue estimation is $|\mathcal{A}|$.*

PROOF. Consider a vehicle v and that all ads are relevant to v. When v enters or leaves the coverage of an RSU, the revenue estimation of all ads has to be updated, thus the complexity is $|\mathcal{A}|$. □

Clearly, when the number of ads and vehicles involved is large, the update procedure becomes cumbersome. To overcome this issue, below we propose an effective approach which greatly reduces the number of ads to consider.

4.2 Sparse-set approximation

An intuitive solution to speed up the ad selection procedure consists in limiting the size of the set of ads $\mathcal{A}$. However, blindly removing ads would wantonly impair the system performance: the problem is not that there are too many ads, but there are too many ads *similar to each other*, hence with the same target audience. We therefore replace the set $\mathcal{A}$ with its *sparse approximation*, as defined below. For the sake of clarity, we start by considering $M = 1$, i.e., each vehicle can display at most one ad per time step.

DEFINITION 4.1 (SPARSE SET). *$\mathcal{X} \subseteq \mathbb{F}$ is a sparse set if, for any two points $\vec{f}_1, \vec{f}_2 \in \mathcal{X}$, $D(\vec{f}_1, \vec{f}_2) > 2\epsilon$.*

It is important to note that, due to local ads that are relevant only to RSUs within their target location, different RSUs may select different ads to be part of their sparse approximation.

The following result states that, given a sparse set of ads $\mathcal{X} \subseteq \mathbb{F}$, the distance between a point in $\mathcal{X}$ and any other point either in $\mathcal{X}$ or in $\mathcal{V}$ ($\mathcal{V} \subseteq \mathbb{F}$), is at least ϵ. It follows that, given $D_{\max}$, a vehicle cannot find in $\mathcal{X}$ more than $\lceil (D_{\max}/\epsilon)^n \rceil$ ads that are relevant to itself.

Theorem 3. *Given a sparse set $\mathcal{X} \subseteq \mathcal{A}$, for every point $\vec{f} \in \mathcal{X} \cup \mathcal{V}$, a closed ball of radius ϵ around $\vec{f}$ contains at most a single ad $a \in \mathcal{X}$.*

PROOF. In the case where $\vec{f} \in \mathcal{X}$ is an ad, the theorem holds given the definition of sparse set. Next, consider that $\vec{f}$ is a vehicle. we prevent the selection of any additional ad within a ball of $2 \cdot \varepsilon$ from a. Assume that there are two ads a_1 and a_2 in $\mathcal{X}$ s.t. $D(a_1, \vec{f}) \leq \epsilon$ and $D(a_2, \vec{f}) \leq \epsilon$. Then, by triangular inequality, $D(a_1, a_2) \leq D(a_1, \vec{f}) + D(a_2, \vec{f}) \leq 2\epsilon$, which contradicts the definition of sparse set. Thus the thesis is proven. □

Let us now introduce the sparse approximation of an ad set.

DEFINITION 4.2 (SPARSE APPROXIMATION). *The sparse approximation of a set of ads $\mathcal{A}$ is a set $\mathcal{A}_\epsilon^{(1)} \subseteq \mathcal{A}$ such that: (i) $\mathcal{A}_\epsilon^{(1)}$ is a sparse set, and (ii) for each ad $a \in \mathcal{A} \setminus \mathcal{A}_\epsilon^{(1)}$, there exists $a' \in \mathcal{A}_\epsilon^{(1)}$ with $r(a') \geq r(a)$ and $D(a, a') \leq 2\epsilon$.*

Intuitively, $\mathcal{A}_\epsilon^{(1)}$ is a sparse set obtained by removing redundant, low-value ads from $\mathcal{A}$. Alg. 2 provides a technique to build the sparse approximation of $\mathcal{A}$.

Algorithm 2 Building a sparse approximation of the ad set (function `EpsilonSet`)

Require: $\mathcal{A}, \epsilon$
1: $\mathcal{A}_\epsilon^{(1)} \leftarrow \emptyset$
2: **sort** $a \in \mathcal{A}$ **by** $r(a)$ in decreasing order
3: **for all** $a \in \mathcal{A}$ **do**
4: $\mathcal{A}_\epsilon^{(1)} \leftarrow \mathcal{A}_\epsilon^{(1)} \cup \{a\}$
5: $\mathcal{A} \leftarrow \mathcal{A} \setminus \{b \in \mathcal{A} \colon D(a,b) \leq 2\epsilon\}$
return $\mathcal{A}_\epsilon^{(1)}$

The algorithm first sorts the ads in the original set by their value (line 2). Then, at each iteration, it adds the top (i.e., highest-value) ad to the sparse set (line 4) and removes all other ads in $\mathcal{A}$ at distance less than or equal to 2ϵ from said ad (line 5). An example of how Alg. 2 works is presented in Fig. 1. It is straightforward to see that, by construction, the resulting set $\mathcal{A}_\epsilon^{(1)}$ is the sparse approximation of $\mathcal{A}$, as by Definition 4.2.

Next, we consider $M > 1$ (i.e., vehicles can display more than one ad per time step). In this case, the broker should select multiple

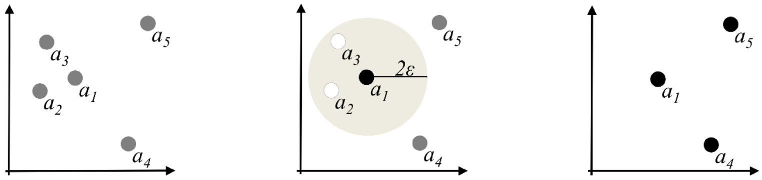

Figure 1: How Alg. 2 works: given the initial set $\mathcal{A}$ with $r(a_1) > r(a_2) > \ldots > r(a_5)$ (left), it first includes a_1 in the sparse set $\mathcal{A}_\epsilon^{(1)}$. Then, a_2 and a_3 are excluded as stated in line 5 (center). Finally, a_4 and a_5 are added to $\mathcal{A}_\epsilon^{(1)}$ (right).

ads targeting the same audience. We therefore introduce the notion of M-sparse set and M-sparse approximation, $\mathcal{A}_\epsilon^{(M)}$.

DEFINITION 4.3 (M-SPARSE SET). *$\mathcal{X}^{(M)} \subseteq \mathbb{F}$ is an M-sparse set if, for any point $\vec{f} \in \mathcal{X}^{(M)}$, there are at most M points within distance 2ϵ from $\vec{f}$ (including $\vec{f}$ itself).*

DEFINITION 4.4 (M-SPARSE APPROXIMATION). *The M-sparse approximation of a set of ads $\mathcal{A}$ is a set $\mathcal{A}_\epsilon^{(M)} \subseteq \mathcal{A}$ such that (i) $\mathcal{A}_\epsilon^{(M)}$ is M-sparse, and (ii) for each subset $\mathcal{B} \subseteq \mathcal{A}$ with $|\mathcal{B}| \leq M$, there exists a subset $\mathcal{B}_\epsilon^{(M)} \subseteq \mathcal{A}_\epsilon^{(M)}$ with $|\mathcal{B}_\epsilon^{(M)}| = |\mathcal{B}|$ and bijection function, $g : \mathcal{B} \to \mathcal{B}_\epsilon^{(M)}$, s.t. $\forall b \in \mathcal{B}$: $r(g(b)) \geq r(b)$ and $D(b, g(b)) \leq 2\epsilon$.*

Algorithm 3 Building the M-sparse approximation of the ad set

Require: $\mathcal{A}, \epsilon, M$

```
A_ε^(0) ← ∅
for j = 1 to M do
    A_ε^(j) ← A_ε^(j−1) ∪ EpsilonSet(A \ A_ε^(j−1), ε)
return A_ε^(M)
```

It is easy to see that, by construction, Alg. 3 builds the M-sparse approximation of the ad set $\mathcal{A}$. Indeed, it repeatedly calls the `EpsilonSet` function defined in Alg. 2. As shown by the following theorem, the sparse set resulting from Alg. 3 includes groups of up to M similar ads that are relevant to the same vehicle.

Theorem 4. *Given an M-sparse set $\mathcal{A}_\epsilon^{(M)}$ output by Alg. 3, for every vehicle $v \in \mathcal{V}$, a closed ball of radius ϵ around v contains at most M ads.*

PROOF. The set $\mathcal{A}_\epsilon^{(M)}$ is generated recursively by forming M sparse sets (as it can be seen in Algorithm 3). Each sparse set satisfies Theorem 3 and, thus, contributes with at most a single ad s.t. $D(a, v) \leq \epsilon$. It follows that the maximum number of ads within a closed ball of radius ϵ, centered in v, is equal to M. □

Replacing the original set of ads $\mathcal{A}$ with its sparse approximation $\mathcal{A}_\epsilon^{(M)}$ makes it possible for the broker to streamline the ad selection procedure. In particular, the estimate of the revenue, $R(a)$, can be updated with limited complexity.

Theorem 5. *When performed on $\mathcal{A}_\epsilon^{(M)}$, the complexity of the revenue estimation update is:* $\min\left\{\lceil\left(\frac{M \cdot D_{\max}}{\epsilon}\right)^n\rceil, |\mathcal{A}_\epsilon^{(M)}|\right\}$.

PROOF. It follows from Theorem 4 that for each vehicle v a closed ball of radius ϵ around v contains at most M ads. Therefore, we are left to consider how many such balls fit into a closed ball of radius $D_{\max}$. The maximum number of balls of radius ϵ that fit in such a volume is: $\lceil\left(\frac{M \cdot D_{\max}}{\epsilon}\right)^n\rceil$. Thus the maximum number of ads within distance $D_{\max}$ from the vehicle is the minimum between such a value and the total number of ads in $\mathcal{A}_\epsilon^{(M)}$. □

Clearly, larger values of ϵ allow a greater reduction of the number of ads, hence a faster ad selection. However, as ϵ grows, $a \in \mathcal{A}$ and its corresponding ad, $a' \in \mathcal{A}_\epsilon^{(M)}$, become less similar. It follows that a' may become not relevant to a certain vehicle (i.e., $D(a', v) > D_{\max}$) while a was (i.e., $D(a, v) \leq D_{\max}$). This means that the opportunities of a selection strategy to pick M relevant ads for a vehicle may diminish when the strategy is applied to $\mathcal{A}_\epsilon^{(M)}$ instead of $\mathcal{A}$. However, in Sec. 5 we show that such a performance loss is negligible even for large values of ϵ, e.g., $\epsilon = D_{\max}/4$.

Finally, we remark that the sparse approximation of the ad set needs to be computed only *once*, while the selection algorithm runs every time a new set of ads to be broadcast has to be identified.

4.3 Multi-RSU, multi-step

The ad selection algorithm can be easily extended to networks comprising multiple RSUs and operating for multiple time steps, such as the one considered in our performance evaluation in Sec. 5.

Specifically, when considering multiple time steps, there is no profit in serving vehicles with the same ad multiple times. Thus, the way the estimated revenues R are computed is enhanced as follows: $R(a, u)$ is increased by $r(a, u)$ only if a has not been broadcast to the vehicle before.

Similarly, we can account for the presence of multiple RSUs, i.e., for the fact that vehicles may have received an ad from some RSU they visited in the past. If a vehicle under the coverage of an RSU, $u \in \mathcal{U}$, has been served ad $a \in \mathcal{A}$ in the past by another RSU, $u' \in \mathcal{U}$, the corresponding $r(a, u)$ value is discounted from the revenue estimation $R(a, u)$. This requires the broker (not the advertisers) to know which RSUs the vehicles visited, a piece of information that can be easily gathered from the beacons cars are required to periodically send and that will be available in next-generation network systems [16]. It does *not* require to know which ads were displayed by cars.

5. PERFORMANCE EVALUATION

We evaluate Volfied using a vehicular trace [17] depicting car mobility in Cologne, Germany. The trace refers to a urban area of 28×32 km^2, and models over 110,000 vehicles, during the course of 8 hours. As shown in Fig. 2, 1,000 RSUs were deployed (black dots in the figure) along the busiest roads and at the center of intersections so that vehicles are under radio coverage for approximately 60% of the time they appear in the trace. The RSU radio range is set to 150 m.

Each vehicle is assigned a five-dimensional feature vector. Feature values are sampled from the normal distribution with a mean of 0.5 and standard deviation of 0.15. Similarly, each ad is assigned a five-dimensional feature vector and a value r, both sampled uniformly in the range $(0, 1)$. 90% of the ads are global, the others are local. Recall that local ads can be displayed only within the coverage area of a specific RSU, which is selected at random.

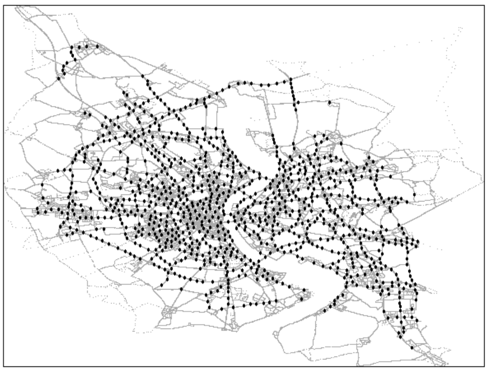

Figure 2: Road layout (grey lines); deployed RSUs are represented by black dots.

We evaluate Volfied, Top-k and a Random strategy by simulating the system over 480 time steps, with each step lasting one minute. At each time step, RSUs broadcast the ads selected by the broker. The tested algorithms only differ from each other in the ad selection strategy, i.e., all of them can access the same information on ads, vehicles and ads values. Specifically, Top-k selects the K ads with highest estimated revenue R, Random selects K random ads among those that have a positive revenue R, and Volfied makes conflict free selection as described in Sec. 4.1. We first perform our evaluation with respect to a *default configuration*, whose settings are: $K = 5$, $M = 1$, $|\mathcal{A}| = 10,000$, $D_{\max} = 0.15$, and $\epsilon = 0.025$, and under the assumption that the presence of vehicles under an RSU can be detected without error. Note that the default configuration implies that the tested algorithms use as input the sparse set $\mathcal{A}^{(1)}_{0.025}$. The performance metrics we plot are:
(i) Total revenue, which reflects the amount of money paid to the broker by advertisers and is computed as the sum of the revenue generated by all broadcast ads (recall that the revenue is equal to the ad value r multiplied by the number of ad impressions);
(ii) Total number of impressions, i.e., the total number of ads that have been displayed by vehicles. This metric reflects the point of view of advertisers who would like to maximize ad visibility;
(iii) The average impression distance, which represents how relevant, on average, a displayed ad is to the user. This last metric clearly accounts for the user's point of view. The lower the distance, the more relevant the displayed ads are to the users. The average distance never exceeds $D_{\max}$.

We remark that other performance metrics such as bandwidth consumption are the same for all tested algorithms and are therefore omitted. Higher performance in any of the metrics we present could also be perceived as more efficient bandwidth utilization.

Table 2: Comparison against the optimum for a single time step

Metric	Top-k	Random	Volfied	Optimum
Revenue	1444.3	810.1	1712.0	1770.3
Impressions	1647	1573	1910	1889
Distance	0.107	0.115	0.125	0.119

Comparison against the optimal solution. We compare the performance of Volfied, as well as that of Top-k and Random, against the optimum derived through (7). To this end, we restrict ourselves to a single-step scenario so that the computation of the optimal solution is viable. The results in Tab. 2 show that Volfied provides near-optimal performance: its revenue is just 3.4% lower than the optimum, while it generates 1% more impressions and similar distance. Note that, since the optimum maximizes the revenue, there may be cases where it selects ads with very high value r but that are displayed by slightly fewer users, while Volfied always generates a conflict-free set thus resulting in a higher number of impressions. The performance gap between the optimum and the other two schemes is much larger: the revenue gain is 25% and 55% when compared to Top-k and Random, respectively.

Performance over time. Fig. 3 shows the time evolution of our performance metrics for the default configuration. As can be observed, Volfied generates 70% higher revenue and 50% more impressions than Top-k. This implies that Volfied satisfies the interests of both broker and advertisers. Because Volfied aims at maximizing the broker's revenue, it may select ads that are slightly (by about 0.01) less relevant to users with respect to Top-k and Random, as shown by the right plot in Fig. 3.

Effect of the ad set size. Fig. 4 shows the impact of the ad set size, $|\mathcal{A}|$, on the system performance. Intuitively, the larger $|\mathcal{A}|$, the easier it is to find relevant ads to each vehicle. Indeed, revenue (and also impressions, omitted for brevity) improves for larger values of $|\mathcal{A}|$. Interestingly, for 1,000 ads the difference between the algorithms is small and Volfied generates 30% more revenue than its alternatives. However, as $|\mathcal{A}|$ increases, the performance gap also grows, and when $|\mathcal{A}| = 20,000$, Volfied increases revenue by 70% with respect to the other schemes. Indeed, the more ads in the system, the more critical their selection becomes and the more severe the revenue loss that occurs due to the conflict discussed in Sec. 3.1. Hence the advantage provided by Volfied becomes more evident.

How many ads to serve? The value of K corresponds to the bandwidth that is consumed by ad broadcasting. The left plot in Fig. 5 shows that, for small values of K, the broker's revenue increases with K. However, it is interesting to notice that the revenue saturates as eventually the vehicles' ability to display ads and the number of possible advertisements become a performance bottleneck. Thus, there is a preferred value of K (which depends on the system settings) that the broker should use.

Furthermore, it is surprising to notice that the performance of Top-k and Random is not monotone with K: increasing K beyond a certain point actually hurts the system performance. The reason is twofold. First, the larger K, the more likely the conflicts. Second, ads that were broadcast before are not considered as profitable anymore (although they can still generate revenue if not all vehicles displayed them); thus, once the top K ads have been broadcast, it becomes increasingly harder to identify the best ads to transmit. Interestingly, Top-k reaches its peak value of revenue for a lower K than Random, due to the fact that the ads selected by Top-k create conflicts more often than those that are randomly chosen. This is confirmed by the right plot in Fig. 5, which shows that the conflicts generated by Top-k reduce the distance between ad and user, hence providing slightly smaller average distance than Random.

Effect of M. We now fix $K = 5$ and study the performance as M varies. Recall that a small value of M accounts for the reduced screen size aboard a vehicle and for the limited driver's attention span, and that typically $M < K$. As shown in Fig. 6, Top-k and Random are very sensitive to M. For $M < K$, they provide much lower revenue and number of impressions; only when M approaches K, i.e., when conflicts seldom occur, Top-k gives good performance. Volfied, instead, is much more robust, as its performance varies very little with M. It generates just 10% lower revenue and 15% fewer impressions when $M = 1$ than when $M = K$.

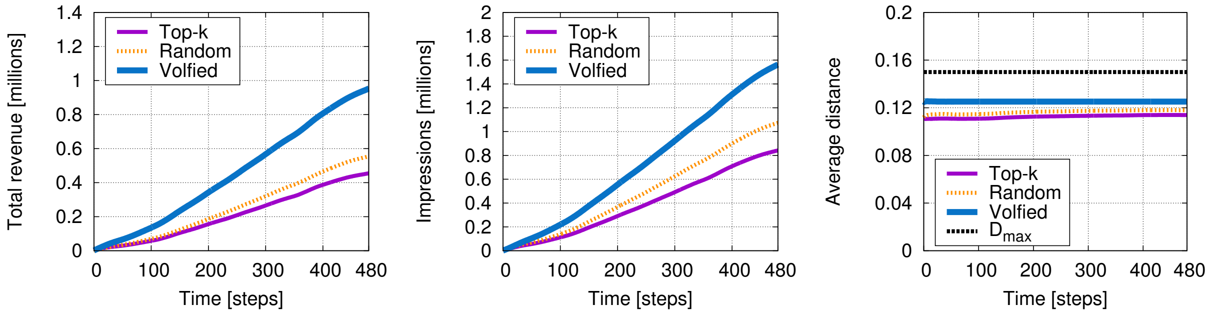

Figure 3: Time evolution of the cumulative revenue, cumulative number of impressions and average distance for Volfied, Top-k and Random (default configuration).

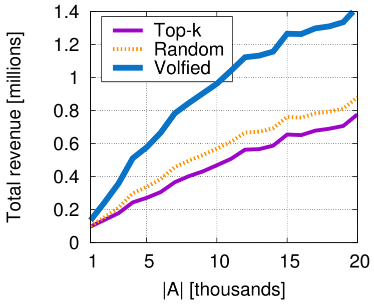

Figure 4: Effect of the ad set size on broker's revenue, when $K = 5$ and $M = 1$.

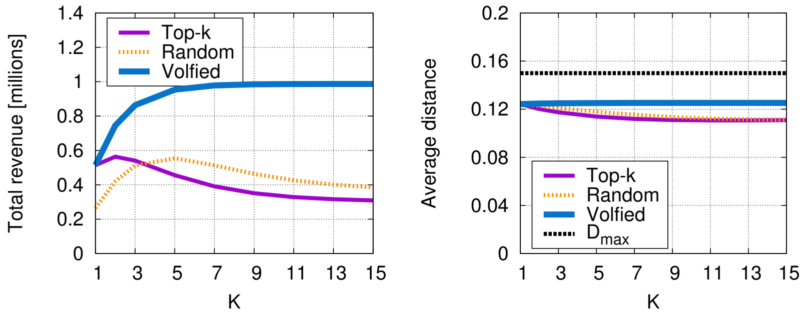

Figure 5: Effect of number of broadcast ads (K) per time step, with fixed $M = 1$.

We remark that the latter is the special case where Volfied and Top-k yield the same revenue, which coincides with the optimum (see also Sec. 4.1). Thus, the fact that Volfied gives a similar revenue for $M = 1$ and for $M = K$, confirms that its performance is near-optimal for any $M < K$.

Effect of ϵ. The left plot in Fig. 7 depicts the broker's revenue as ϵ varies, when $D_{\max} = 0.15$. As can be observed, for values of $\epsilon \leq D_{\max}/4$, the revenue loss due to the sparse approximation is negligible. Also, for such values of ϵ we limit the number of ads that have to be processed per vehicle arrival/departure. As shown in Theorem 5, the number of processed ads is bounded by: $\lceil \left(\frac{D_{\max}}{\epsilon}\right)^n \rceil$ regardless of the size of the ad set $\mathcal{A}$, which, for $\epsilon = D_{\max}/4$, amounts to $4^5 = 1024$.

Vehicle detection accuracy. While deriving the previous results, we assumed that an RSU could reliably detect all vehicles under its coverage thanks to their beacon messages. The right plot in Fig. 7 shows the impact of different levels of accuracy, i.e., probability of successfully detecting a vehicle under an RSU. Remarkably, Volfied with 0.3 accuracy provides higher revenue than the best alternative with accuracy equal to 1. It follows that Volfied is very effective even with incomplete knowledge of the scenario, since it can still successfully avoid conflicts.

$D_{\max}$ and average distance. $D_{\max}$ is another important parameter as it determines which ads are relevant to a user. Intuitively, the larger $D_{\max}$, the easier it is to select relevant ads that will be displayed by a vehicle, but also the larger the average ad-user distance. Fig. 8 confirms these trends for all selection strategies. However, we can see that, when $D_{\max}$ is very small, all strategies yield similar revenue and average distance as the set of ads with positive revenue, hence that can be selected, is very small. Likely each vehicle has at most one ad within distance $D_{\max}$ in $\mathcal{A}_\epsilon^{(1)}$. For larger values of $D_{\max}$, instead, Volfied provides higher revenue than the other schemes, as conflicts become increasingly likely and cause revenue loss (left plot in Fig. 8). For Volfied, the price to pay is a slight increase in the average ad-user distance (right plot in Fig. 8).

6. CONCLUSIONS

We addressed targeted advertising in vehicular networks and envisioned a system where advertisers pay a broker based on the value and the number of impressions of each ad. We considered the broker's perspective and formulated the problem of selecting the ads to broadcast that maximize the broker's revenue, subject to a maximum number of ads that can be transmitted. While doing this, we identified a conflict between user and broker's interests, which severely hurts the broker's revenue if not properly addressed. Then, in light of the problem complexity, we introduced Volfied, an efficient greedy algorithm that always selects a conflict-free set of ads while maximizing the broker's revenue. The complexity of Volfied has been proved to be linear with the number of ads. In addition, we proposed a sparse approximation of the ad set, which further speeds up ad selection. We evaluated Volfied and our sparse approximation technique in a realistic vehicular environment, against the optimum in a single-time step scenario and against the Top-k and Random strategies in a multi-time step scenario. Our results show that Volfied provides near-optimal performance. Also, it improves the broker's revenue by up to 70%, and the number of displayed ads by up to 50%, with respect to Top-k.

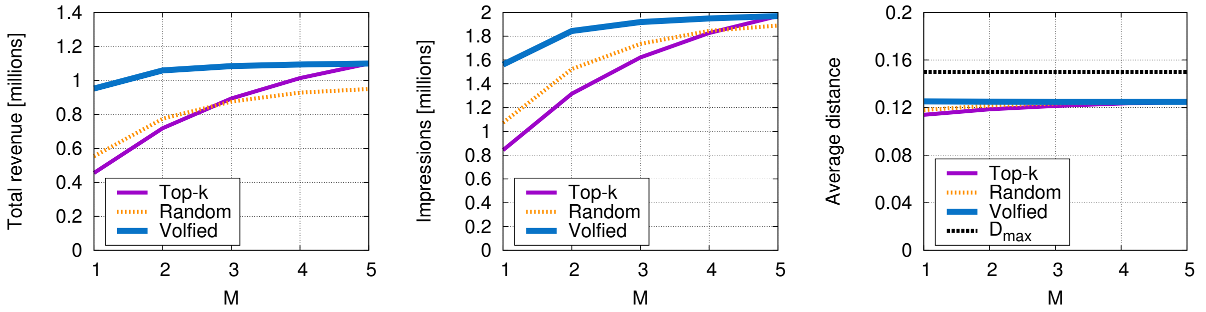

Figure 6: Effect of the number of displayable ads (M) on performance metrics for fixed $K = 5$.

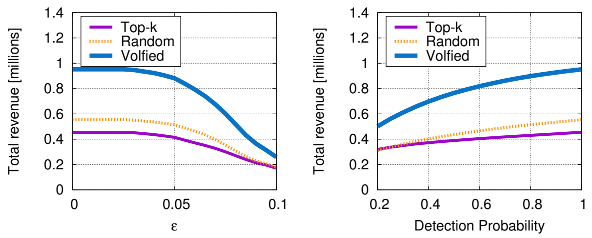

Figure 7: Effect of ϵ (left) and of accuracy in vehicle detection (right) on the broker's revenue.

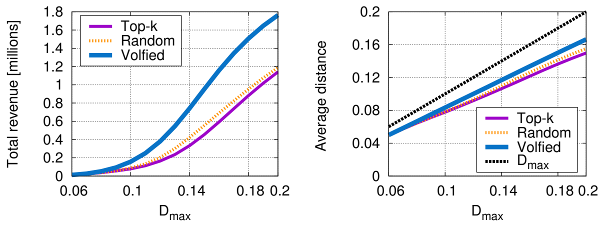

Figure 8: Effect of $D_{\max}$ on performance metrics.

7. ACKNOWLEDGMENT

This paper was made possible by NPRP grant $\sharp\, 5-782-2-322$ from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

8. REFERENCES

[1] G. Sterling, "It's official: Google says more searches now on mobile than on desktop," *SearchEngineLand*, 2015.

[2] R. van der Meulen and J. Rivera, "Gartner says mobile advertising spending will reach 18 billion dollars in 2014," 2014, Gartner Press Release.

[3] A. Rao, F. Schaub, and N. M. Sadeh, "What do they know about me? Contents and concerns of online behavioral profiles," *CoRR*, vol. abs/1506.01675, 2015.

[4] H. Beales, "The value of behavioral targeting," *Network Advertising Initiative*, Jan. 2010.

[5] W. Wang, L. Yang, Y. Chen, and Q. Zhang, "A privacy-aware framework for targeted advertising," *Computer Networks*, vol. 79, pp. 17–29, Mar. 2015.

[6] Z. Abbassi, A. Bhaskara, and V. Misra, "Optimizing display advertising in online social networks," in *ACM WWW*, 2015.

[7] J. Feldman, N. Korula, V. Mirrokni, S. Muthukrishnan, and M. Pál, "Online ad assignment with free disposal," in *Internet and Network Economics*, ser. Lecture Notes in Computer Science, S. Leonardi, Ed. Springer Berlin Heidelberg, 2009, vol. 5929, pp. 374–385.

[8] R. Friedman and A. Libov, "An advertising mechanism for p2p networks," in *IEEE P2P*, 2013.

[9] G. Aust, *Vertical Cooperative Advertising and Pricing Decisions in a Manufacturer-Retailer Supply Chain: A Game-Theoretic Approach*. Cham: Springer International Publishing, 2015, pp. 65–99.

[10] A. Reiffers-Masson, E. Altman, and Y. Hayel, "A time and space routing game model applied to visibility competition on online social networks," in *NETGCOOP*, 2014.

[11] A. Reiffers-Masson, Y. Hayel, and E. Altman, "Game theory approach for modeling competition over visibility on social networks," in *COMSNETS*, 2014.

[12] E. Altman and N. Shimkin, "Strategic posting times over a shared publication medium," in *NETGCOOP*, 2014.

[13] F. P. Kelly, P. Key, and N. S. Walton, "Incentivized optimal advert assignment via utility decomposition," *ACM Conference on Economics and Computation*, 2014.

[14] E. Vee, S. Vassilvitskii, and J. Shanmugasundaram, "Optimal online assignment with forecasts," in *ACM Conference on Electronic Commerce*, 2010.

[15] C. Borgiattino, C.-F. Chiasserini, F. Malandrino, and M. Sereno, "Advertisement delivery and display in vehicular networks," in *IEEE VTC-Fall*, 2015.

[16] E. T. P. for Communications Networks and Services, http://networld2020.eu/wp-content/uploads/2015/01/Joint-Whitepaper-V12-clean-after-consultation.pdf.

[17] D. Naboulsi and M. Fiore, "On the instantaneous topology of a large-scale urban vehicular network: The Cologne case," in *ACM MobiHoc*, 2013.